CALS TEST NETWORK

CTN Test Report
91-020

UCRL-ID-110239

Raster Data Transfer Test Using DATA DEVELOPMENT Inc. Produced Data: MIL-R-28002 Type I Raster

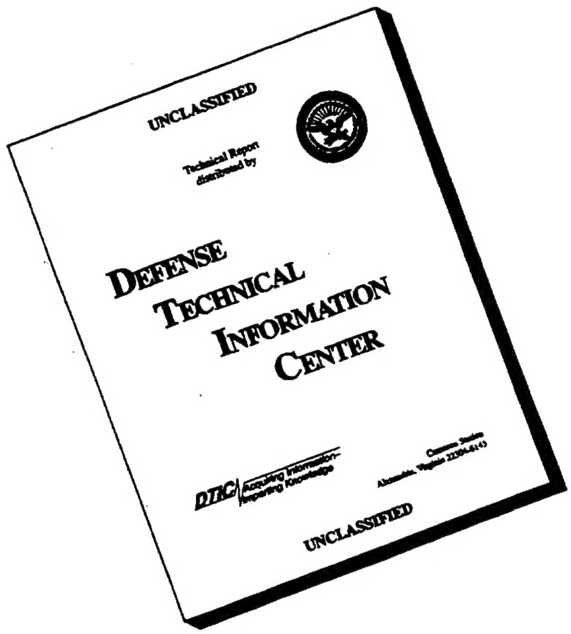
Quick Short Test Report

March 18, 1992

Prepared for
Air Force Materiel Command

Prepared by
Lawrence Livermore National Laboratory

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

CTN Test Report
91-020

UCRL-ID-110239

Raster Data Transfer Test Using DATA DEVELOPMENT Inc. Produced Data: MIL-R-28002 Type I Raster

Quick Short Test Report

March 18, 1992

Prepared by
Lawrence Livermore
National Laboratory

LLNL Contact
Nick Mitschkowetz
(510) 422-0582

Donald L. Vickers
(510) 422-4231

CTN **Contact**
Mel Lammers
(513) 257-3085

Prepared for
Air Force Materiel Command
AITI Project
CALS Test Network (AFMC/ENCT)
Wright-Patterson AFB, OH 45433-5000

Prepared by
Lawrence Livermore National Laboratory
Livermore, CA 94550

This report has been reproduced
directly from the best available copy.

Available to DOE and DOE contractors from the
Office of Scientific and Technical Information
P.O. Box 62, Oak Ridge, TN 37831
Prices available from (615) 576-8401, FTS 626-8401

Available to the public from the
National Technical Information Service
U.S. Department of Commerce
5285 Port Royal Rd.,
Springfield, VA 22161

Contents

APPENDIX

1 Introduction

1.1 Background and Test Objectives

The DoD Computer-aided Acquisition and Logistic Support (CALS) Test Network (CTN) is conducting tests of the military standard for the Automated Interchange of Technical Information, MIL-STD-1840A, and its companion suite of specifications. The CTN is a DoD-sponsored confederation of voluntary participants from industry and government, managed by the Air Force Logistics Command.

The primary purpose of the CTN is to evaluate the effectiveness of the CALS standards for technical data interchange and to demonstrate the capability and operational suitability of these standards.

To this end, testing should represent the systems and applications in use by a large number of participants. Sampling a wide cross-section of industry and government will gain feedback on the various interpretations of the standards and broaden the base of industry participation in the CALS initiative.

This test was conducted to allow DATA DEVELOPMENT, Inc., to demonstrate their ability to generate a MIL-R-28002 data file. The objective was to evaluate their interpretation of the MIL-R-28002 standards, thereby assisting the CTN in substantiating the validity of the standards or recommending changes to these standards and the references to national or international standards.

Additionally, Quick Short Test Reports (QSTRs) are intended to promote industry and government participation in the CALS initiative, developing a level of confidence in the technology, and furthering mission objectives.

1.2 Purpose

The following test was undertaken to develop a more comprehensive understanding of the procedural issues and data structure issues surrounding a conference-floor data interchange undertaken by DATA DEVELOPMENT and the CTN at the CALS EXPO '90 in Dallas, Texas.

Previous experiences with CALS data outside the CTN arena suggested that DATA DEVELOPMENT was CALS ready. A sample technical publications tape generated by DATA DEVELOPMENT was brought over to the CTN booth at CALS EXPO '90 for a "walk-in" evaluation.

The "walk-in" test produced unexpected results. For reasons not immediately apparent, the majority of the data were not being transferred successfully. Since the hardware platform and the show environment was somewhat less than optimum for conducting an in-depth test, it was decided that a more comprehensive test should be undertaken on the CTN Raster Test Bed at the CTNO/LLNL.

Since the CALS initiative is intended to develop a system independent data interchange strategy, a very important aspect of the test was to determine the reason why some systems were apparently able to interpret the transfer data and other systems were not.

2 Test Parameters

Date:

4-February-1991

Test Plan:

Information is to be transferred to the CTNO/LLNL Raster Test Bed on a MIL-STD-1840A tape. The first tape tested will be one of the tapes presented to the CTN at CALS EXPO '90. The CTN shall retain custody of the tape at CALS EXPO and transport the tape to the Raster Test Bed for further analysis.

The CTN shall parse the tape and capture all the data available. However, for the purposes of this QSTR, only the tape format and a portion of the Raster data on the tape shall be evaluated.

By mutual agreement, the CTN may request DATA DEVELOPMENT to provide additional data on a separate tape.

Evaluators:

Lawrence Livermore National Laboratory
P.O. Box 808, L-542
Livermore, CA 94551

Data Originator:

DATA DEVELOPMENT Inc.
49 West Flagler Ave.
Stuart, FL 34994

Data Description:

MIL-STD-1840A tape containing Technical Publication data, including text and raster graphics files with the appropriate DTD information.

Data Source System:

Evaluation Tools Used:

Sun 3/280 (UNIX)

TAPETOOL	CTN tape evaluation routine
CALSTB.350	CTN/CALS raster utility
OD	UNIX octal dump utility
ALTER	CTN file edit utility

DEC Micro VAX-II (VMS)

TAPETOOL	CTN tape evaluation routine
VALIDG4	SYSCON group-4 evaluator
DUMP	DEC octal dump utility

Standards Tested:

MIL-STD-1840A
MIL-R-28002A

3 MIL-STD-1840A Analysis

3.1 1840A External Packaging Analysis

Analysis of the external packaging, in this test, is not applicable. The data was hand-delivered to the CTN at the CALS EXPO '90 and hand-carried back to the CTNO/LLNL for analysis.

3.2 1840A Transmission Envelope Analysis

3.2.1 Tape Formats

3.2.1.1 Sun/UNIX TAPETOOL Evaluation

This evaluation determined that the tape format conformed to FIPS PUB 25, 50, 79 (see 1840A 5.2.1). However, a data file block length anomaly was flagged by TAPETOOL for every MIL-R-28002 image data file. The error indicated that the last data block in the file was less than 2048 bytes long (see 1840A 5.2.1.6).

3.2.1.2 DEC/VMS TAPETOOL Evaluation

This evaluation also determined that the tape format conformed to FIPS PUB 25, 50, 79 (see 1840A 5.2.1). However, unlike the Sun/UNIX evaluation, reading the same tape on the MicroVAX generated no block length errors.

Both TAPETOOL applications (DEC and Sun) were written in "C". The TAPETOOL strategy is to use the highest level system's utilities possible, thus avoiding low level system input/output (I/O) calls. This provides a more uniform usage of the hardware platforms hosting it.

Presumably, this approach reflects the same strategy that would be employed by a programmer tasked to implement a CALS utility in the most expedient manner, allowing the system to handle the lower level data transfer issues such as file formats, record structure, etc.

To examine the actual data on the MIL-STD-1840A tape, the tape contents were dumped directly to the system's console. Using the DEC/VMS system "dump" utility, the tape format and data were viewed in both octal and ASCII. The dump indicated that the actual block lengths of the MIL-R-28002 files were correct. There were no short blocks at the end of each raster image. These blocks were all padded out to the full 2048-byte length with the circumflex-accent character ("^").

The I/O utilities being used by the Sun/UNIX version of TAPETOOL were obviously sensitive to that padding and were skipping over it on input. The circumflex accent, as a padding character, is required in MIL-R-28002 as specified by ANSI X3.27 (see ANSI X3.27 6.3.4).

The I/O utilities being used by the Sun/UNIX version of ANSITAPE were not sensitive to the circumflex accent character. However, this I/O process is sensitive to a system specific flag in the ANSI X3.27 tape header HDR2. Refer to the Appendix-A- for the documentation on the "ANSITAPE anomaly".

3.2.2 File Formats

All Analysis indicated that the tape contained ten (10) documents with as many Declaration files (D001 through D010). All the data files indicated in the Declaration files were present.

A file numbering scheme anomaly was flagged by TAPETOOL. It indicated that the file numbering sequence within each of the document sets was not contiguous (see 1840A 5.1.3).

Despite this anomaly, all the files in the data transfer were uniquely labeled in a fashion closely aligned with MIL-STD-1840A.

3.2.2.1 Declaration Files

No Declaration file anomalies were encountered. All required Declaration file records were presented in the test data. The files were all variable length ANSI type-"D" files (see 1840A 5.1.1.1).

3.2.2.2 Data Files (MIL-R-28002)

Several mechanisms were used to evaluate the interchange data. Each evaluation mechanism used a slightly different strategy for employing both Host system's utilities and applications codes to capture the data from the transfer media (MIL-STD-1840A 9-track magnetic tape).

In the Sun/UNIX environment, both the CTN TAPETOOL utility and the public domain ANSITAPE were used to access MIL-R-28002 data files. In the VAX/VMS environment, only the CTN TAPETOOL utility was used.

The object of this test strategy was to determine the reason for the apparent variation in data interchange results that had been experienced over a range of dissimilar hardware platforms using (reportedly) the same data tape.

3.2.2.2.1 Sun/UNIX TAPETOOL Evaluation:

Using the CTN TAPETOOL in the Sun/UNIX system, all the indicated tape files were successfully read from tape to disk. However, none of the raster image (MIL-R-28002) files could be viewed.

Using the UNIX octal dump utility, OD, it was determined that the CCITT Group-4 data in all the tested image files started after the last byte of record-11, labeled "notes: ". Instead of starting the binary compressed image data at the beginning of the second data block (byte 2048), the Group-4 data was always starting at byte 1408. The resulting disk file appeared to be missing the block padding required to fill the header out to a full 2048 bytes (see 1840A 5.2.1.6).

A CALS utility expecting to retrieve Group-4 image data from this file would go to byte 2048 and begin to decode what was thought to be the start of the image. In actuality, the data being retrieved was some distance into the compressed binary information. Naturally, the resulting decompression was always erroneous.

3.2.2.2.2 Sun/UNIX ANSITAPE Evaluation:

Using ANSITAPE on the Sun/UNIX system, all the indicated tape files were successfully read from tape to disk. Again, none of the MIL-R-28002 image files could be viewed in their current disk structure.

This time, the UNIX octal dump utility revealed that at the end of every 128-byte fixed-length record (see 1840A 5.2.1.6) there occurred a "line-feed" (octal 12) character. Additionally, each "line-feed"

character was followed by a "circumflex-accent" (octal 136). The fixed-length records in all the MIL-R-28002 image files had been so altered. The effective increase in record length to 129 bytes pushed the header data block length out to 2064 bytes.

A CALS utility expecting to retrieve Group-4 image data from this file would go to byte 2048 and begin to decode what was thought to be the start of the image. In actuality, the data being retrieved was the overflow padding character from the header. Again, the resulting decompressions were always erroneous.

3.2.2.2.3 DEC/VMS TAPETOOL Evaluation

Using the CTN TAPETOOL on the DEC/VAX system, all the indicated tape files were successfully read from tape to disk. In this environment, several of the images were successfully displayed. The first sixteen (16) MIL-R-28002 files were separated into header and Group-4 data. Using the image parameters found in the header, the related Group-4 encoding was evaluated using the CTN VALIDG4 utility. All images successfully passed the evaluation test.

3.2.2.2.4 Header Data

No header data errors were encountered through automated analysis by either the DEC or Sun systems. The headers of the image files selected for review were all correct as required.

3.2.2.2.5 Image Data Evaluation Procedure

To determine the validity of the image data contained on the MIL-STD-1840A tape, the MIL-R-28002 image files read from the tape by both the Sun/UNIX system and the VAX/VMS systems were analyzed.

The MIL-R-28002 files copied from tape to the Sun, although containing all the appropriate data, had not been copied onto the Sun in a displayable form. The first image file ("D001R001") was selected for reprocessing. Using ALTER, the irregular sized header was stripped from the Group-4 binary data. CALSTB.350 was used to generate a correct CALS header and re-attach to the Group-4 binary data. The resulting MIL-R-28002 file was then successfully displayed on the Sun using CALSTB.350.

Since the VAX had read the MIL-R-28002 files from tape without error, the VAX files were chosen to evaluate the CCITT Group-4 data.

4 MIL-R-28002 Raster Image Analysis

The first sixteen (16) image files encountered on the tape were selected as a representative sampling for Group-4 analysis using the CTN VALDIG4 utility. The header parameters and the Group-4 data were passed to VALIDG4 for encoding evaluation. All the sample files passed this test. The first two raster files on the tape ("D001R001" and "D002R001") were selected for further raster analysis to determine if the subsequent scan lines were valid two dimensional compressions.

4.1 File "D001R001" Analysis

The following sections contain computer output from the analysis of raster image file "D001R001".

4.1.1 File Header Record Data

These are data from the header.

```
srcdocid:   JC074000
dstdocid:   PSDS Conversion
txtfilid:   W
figid:      0001
srcgph:     JCO74A01
doccls:     Unclassified
rtype:      1
rorient:    000,270
rpelcnt:    004320,001785
rdensty:    0600
notes:      BDM-1292            ORG
```

4.1.2 File Structure

This section contains file structure data from both MIL-R-28002 and CCITT analysis.

4.1.2.1 MIL-R-28002 Evaluation

These are image attributes.

```
File size:                31872 bytes
Header Block Size 2048:   2048 bytes
Record Size 128:          128 bytes
Block Padding:            "^" character
```

4.1.2.2 CCITT Recommendation T.6 Evaluation

This is the CCITT Group-4 encoding.
First line decoding: Valid T.6 Group-4 encoding.

```
177777 177777 177777
blank scan lines

            177440 037331 121051
            1111111100100000 0011111011011001 1010001000101001
            vert(0) x8
                    001
                      hor.
                       00000 0011111
                             make-up(2560)
                                    011011001
                                      make-up-white(1280)
                                              10100
                                                   term.-white(9)
                                                         010
                                                            term.-black(1)
                                                           001
                                                           hor.
                                                         01001 11
                                                            term.-white(18)


            144355 055073 044214
            1100100011101101 0101101000111011 0100100010001100
              0010
                term.-black(6)
                  001
                  hor.
                     11011
                          make-up-white(64)
                          01 01011
                               term.-white(25)
                                  010
                                      term.-black(1)
                                     001
                                      hor.
                                        11011
                                           make-up-white(64)
                                              0100100
                                                    term.-white(27)
                                                     010
                                                      term.-black(1)
                                                        001
                                                          hor.
```

```
           062064 057440 042153
100 0110010000110100 0101111100100000 0100010001101011
           0
100 0
term.-white(3)
     11
     term.-black(2)
       001
       hor.
          000011
          term.-white(13)
                010
                term.-black(1)
                   0 01
                   hor.
                       0111
                           term.-white(2)
                           11
                           term.-black(2)
                             001
                             hor.
                                0000 010
                                 term.white(29)
                                         0010
                                         term.-black(6)
                                             001
                                             hor.
                                                101011
                                                term.-white(17)
```

```
043314 160623 114662
0100011011001100 1110000110010011 1001100110110010
010
term.-black(1)
   001
   hor.
      1011
      term.-white(4)
          0011
          term.-black(5)
              00 1
              hor.
                  1100
                  term.-white(5)
                      0011
                      term.-black(5)
                          001
                          hor.
                             0011
                             ter.-white(5)
                                  10
                                  term.-black(3)
                                    011
                                    vert(1)right
-----new line----
                                       001
                                       hor.
                                          1011
                                          term.-white(4)
```

4.1.3 Image Presentation Evaluation (MIL-R-28002)

This section contains visual evaluations of the image quality using the CTN Technical Transfer Procedures Checklist as a guideline.

Decompression and Display:

The file decompressed without encoding errors; the image displayed was that of a table of text. The image was right-reading, commensurate with the header orientation parameter.

Cropped and Centered:

The cropping was close. The image showed a relatively even 56 pel margin of white space around its border, leaving it nicely centered.

Orthographic Alignment:

The thick black border outline was reasonably parallel to the CTR display format on all sides. No skew or parallelogram shift was noted.

Image Continuity:

The image appeared to be complete with no obvious dropouts or misalignments in the image artifacts.

Image Fidelity:

The image showed excellent contrast and there was no perceivable background "noise" present.

5 MIL-R-28002 Raster Image Analysis

5.1 File "D002R001" Analysis

The following sections contain computer output from the analysis of raster image file "D002R001".

5.1.1 File Header Record Data

These are data from the header.

```
srcdocid:    JC084000
dstdocid:    PSDS Conversion
txtfilid:    W
figid:       0001
srcgph:      JC084A01
doccls:      Unclassified
rtype:       1
rorient:     000, 270
rpelcnt:     002032, 002718
rdensty:     0600
notes:       BDM-1706              ORG
```

5.1.2 File Structure

This section contains file structure data from both MIL-R-28002 and CCITT analysis.

5.1.2.1 MIL-R-28002 Evaluation

```
These are image attributes.
File size:                  20352 bytes
Header Block Size 2048:          2048
Record Size 128:            128
Block Padding:              "^" character
```

5.1.2.2 CCITT Recommendation T.6 Evaluation

This is the CCITT Group-4 encoding.
First line decoding: Valid T.6 Group-4 encoding.

```
177777 177777 171056
blank lines at top of page (32)
                   1111001000101110
                   vert(0) x4
                           001
                           hor.
                          00010111
                          term.-white(21)

      031603 027144 160240
    0 0011001110000011 0010111001100100 1110000010100000
    0 0011
    term.-black(7)
          001
          hor.
             1100
             term.-white(5)
                 00011
                 term.-black(7)
                       001
                       hor.
                          0111
                          term.-white(2)
                              0011
                              term.-black(5)
                                  001
                                  hor.
                                     00 111
                                     term.-white(10)
                                           00000101000
                                           term.-black(23)

      170140 171063 147663
    00 1111000001100000 1111001000110011 1100111110110011
    00 1
    hor.
        1110
        term.-white(6)
            000011000
            term.-black(15)
                     00 1
                     hor.
                        1110
                        term.-white(6)
                            010
                            term.-black(1)
                               001
                               hor.
                                  10011
                                  term.-white(8)
                                        11
                                        term.-black(2)
                                          001
                                          hor.
```

```
   114744 076422 132410
1 1001100111100100 0111110100010010 1011010100001000
1 100
term.-white(5)
     11
     term.-black(2)
       001
       hor.
          1110
          term.-white(6)
              010
              term.-black(1)
                 0 01
                 hor.
                     1111
                     term.-white(7)
                         010
                         term.-black(1)
                            001
                            hor.
                               0010 1011
                               term.-white(42)
                                        010
                                        term.-black(1)
                                           1
                                           vert(0)
---new line----
                                            000010
                                            vert-left(2)
```

5.1.3 Image Presentation Evaluation (MIL-R-28002)

Decompression and Display:

The file decompressed without encoding errors; the image displayed was that of a flow diagram. The image was right-reading, commensurate with the header orientation parameter.

Cropped and Centered:

The cropping was close. The image showed a relatively even 36 pel margin of white space around its border, leaving it nicely centered.

Orthographic Alignment:

The thick black border outline was reasonably parallel to the CRT display format on all sides. No skew or parallelogram shift was noted.

Image Continuity:

The image appeared to be complete with no obvious dropouts or misalignments in the image artifacts.

Image Fidelity:
The image showed excellent contrast and there was no perceivable background "noise" present. The quality of the original hard-copy did not seem to warrant the high scan resolution used to capture the image. At 600 lines per inch, the scanned image showed all the imperfections of the original.

6 Conclusions and Recommendations

This data interchange test indicated that the magnetic tape was generally in accord with MIL-STD-1840A and the raster image data was in accord with the MIL-R-28002 Type-I specification.

The anomalies encountered at the CALS EXPO '90 and later at the CTNO/LLNL Raster Test reflect the functional differences of the various CALS implementation strategies used to read the data.

Observed Anomalies:

1. Circumflex accent "^" tape block padding caused some CALS implementations to lose synchronization with the start of the Group-4 image data.

2. The CALS Raster Test Bed implementation using the Army ANSITAPE utility inserts erroneous carriage control characters at the end of fixed length records when the appropriate ANSI X3.27 file header flags are not set (see Appendix A).

3. CALS file labeling conventions were not strictly adhered; however, all the files were uniquely identified and related to the appropriate data set by file name structure.

Issues raised by the test:

1. Various CALS implementation may read the same MIL-STD-1840A tape and interpret its content differently.

2. The difference in data interpretation is not necessarily a function of any particular computer platform; it's a function of the strategy and tools (hardware and software) used to implement a particular CALS application.

3. The CALS standards only specify the interchange media and, more importantly, the structure of the data during the transfer. The implementor is free to choose the mechanism by which the media is read and how the data is handled once it resides in the applications environment. However, the implementor is responsible to assure that the data being imported is not inadvertently modified.

4. Two of the three CALS implementations at the CTNO/LLNL inadvertently modified CALS data imported from the Data Development tape. One implementation (TAPETOOL on the Sun) rejected Circumflex accent characters "^" as NUL data. The other implementation (ANSITAPE on a Sun) inserted carriage control characters in response to byte 27 of the "Reserved for Systems Use" filed in ANSI X3.27 header 2 (HDR2).

5. CALS specifies file labeling conventions to assure unique file names and establish a relationship between files that is applications independent. The intent is to provide the same logical relationship to the digital data interchange that segregated aperture card decks provide in conventional technical data interchanges.

It is the intention of CALS to develop a strategy that is applicable over the widest range of computer system environments. By applying well-established industry standards, CALS anticipates most vendors will find the requirements acceptable
and relatively easy to implement.

Implementation strategies will vary widely. Obviously, the more general the strategy, the shorter the implementation time. Generic implementations will rely heavily on higher level systems I/O

utilities. Higher level I/O utilities often don't provide the programmer with the control of the data that is available through the lower level I/O access. However, lower level systems I/O strategies take longer to develop, are not as transportable, and are more difficult to maintain.

In the final analysis, it is the programmer's responsibility to assure that his application interprets the CALS data accurately. The programmer must determine what level of I/O utilities are to be used to read the tape data and to determine how those utilities affect the CALS data formats.

However, some simple strategies may be applied to ease the complexity of implementing an application.

Recommendation:

ANSI X3.27 specifies that circumflex accent ("^") characters be used to pad out unused buffer space. The fact that some systems interpret this character as "NUL" can make it difficult for an application to find the beginning of the binary image.

A simple solution would be to require a MIL-R-28002 file to be filled with data, eliminating the need for padding. All the remaining header records after the 11 records currently required by MIL-R-28002A, should be filled with "space".

This effectively specifies a "space" character padding without conflicting with the ANSI X3.27 standard. This strategy would also allow additional header records to be implemented in the future.

Although requiring the sending system to generate a block of useless data and requiring the receiving system to store unnecessary data, this alternative would fix the starting location of the Group-4 image data.

Developing a consistent file ID strategy is an important requirement. Each file must be uniquely identifiable and its ID must associate it with a particular document in a record set. Within that document, the file should retain a sequential position.

MIL-STD-1840A requires the sequence component of the numbering strategy to be contiguous throughout each document (see 1840A 5.1.3).

A significant number of vendors have implemented the numbering strategy such that, within each document set, all the files of the same type are contiguously numbered.

Because the declaration file ID is a component of the numbering strategy, either scheme provides unique file identifiers for all files on the interchange media.

It appears that it is more convenient, from the implementor's perspective, to process the file types separately within a document. Certainly it is much easier to audit the various file types in one or more documents when the number system is contiguous through file types.

Recommendation:

Having conducted several MIL-STD-1840A document file audits, the CTN has found that having the files of a given type sequentially labeled within a document is a far more useful strategy from the recipient's standpoint.

Additionally, in preparing CTN test data, being able to handle individual file groups contiguously was also much more efficient.

MIL-STD-1840A should be modified to read:
"The first data file, of every file type, in a document shall use "001" and the number shall increment sequentially for each file of that type in the document, so that each file has a unique file name".

Not unlike aperture cards in a draw or shipping container, CALS data files should be written to tape by contiguous set numbers, allowing an application to read off a complete data set without having to search the entire interchange tape for its constituents. Declaration files should be located at the beginning of the tape to allow the application to identify which data sets the tape contains.

APPENDIX A Utility "ansitape.c" Anomaly

UTILITY "ansitape.c" ANOMALY

EXECUTIVE SUMMARY

As the result of a larger number of CALS tests and some intensive "sleuthing," the CTN has uncovered a data format anomaly that profoundly affects CALS data transfers. The discovery of the anomaly cannot be attributed to a single test or series of tests; therefore, the documentation of the anomaly and a recommendation for a strategy to avoid it are presented in this separate CTN report.

During the past 24 months of CTN testing , we have noticed that occasionally a tape generated on one system would produce erroneous raster files on a receiving system. Yet, careful analysis of the tape showed the raster files to be correct.

Recently, we discovered that the problem could be linked to some incorrect assumptions by system implementors about "ansitape.c," a utility commonly used throughout the DoD community for reading and writing 9-track magnetic tapes. The utility was written by the Army and, because of its great value and ease of use, quickly became the "tool-of-choice" for handling ANSI X3.27 formatted magnetic tape media.

This report gives a narrative account of the discovery of the problem, its nature, and our recommendations. Briefly, the problem stems from the use, by certain systems, of a format control parameter (flag) situated in the "Reserved for Systems Use" file of ANSI X3.27 Header 2 (HDR2). Our initial conclusion was to address the issue by referencing the flag in the appropriate CALS standard (MIL-STD-1840A).

However, further evaluation and experience dictate that the format of disk files, generated by any program reading CALS media, is the domain of that application, not the CALS data interchange strategy. Hence, it is the responsibility of the analyst designing a CALS application to develop and integrate utilities that preserve the fidelity of the CALS data for presentation and interchange activities.

It is a part of the CTN charter to support the CALS initiative by publishing observed anomalies in data formats caused by the integration of commonly used utilities. Only those anomalies that alter CALS interchange data shall be addressed by standards modification recommendations.

INTRODUCTION

During an evaluation of an Autotrol-generated Technical Document, transferred to us on a MIL-STD-1840A tape, anomalies surfaced which made the binary files (D001R001 D001R002 and D001R003) unusable.

The problem apparently stems from various implementations of the ANSI magnetic tape standard (ANSI X3.27 1978) where a non-standard "Format Control Flag" in ANSI header number 2 (HDR2) is being "interpreted" by some vendors and the test bed utilities, and ignored by others.

SYMPTOMS

A MIL-STD-1840A tape was parsed by the CTN VAX at LLNL to assess structure, format, and content of the CALS data it contained. The IGES files were evaluated locally (on the VAX). The raster data was networked (TCP/IP) to a SUN/UNIX platform for evaluation.

Of the three files (D001R001 D001R002 D001R003), one decompressed and displayed nicely (D001R002), while the other two were flawed early in their group-4 encoding. Decompression of the flawed files was partially successful but only the first few scan lines of each file could be displayed.

Worried that the flaws may have been introduced by the rather circuitous network route the files had taken, it was decided to load the source tape directly on the SUN using the public domain utility "ansitape.c". The resulting SUN disk files were not usable.

Each Group-4 file (D001R001 D001R002 D001R003) consisted of 129-byte fixed length record, each record having an extra "line-feed" (012 octal) added onto the end. These files were unlike those originally transferred from the VAX to the SUN over the network. However, both sets of files (VAX and SUN) originated from the same tape.

TESTING

The original data tape was dumped to determine if the extra line-feeds were in the raster files on that media; they were not. It was apparent that the VAX was interpreting the tape files differently than the SUN.

Data tapes (CTN preliminary raster test tapes) previously generated by the CTN VAX were read onto the SUN and displayed correctly. Disk files on the SUN, with the anomaly, were read onto tape and transferred to the VAX where "tapetool" evaluated them correctly. It seemed the anomaly was transparent to the VAX but fatal to the SUN.

However, we were finally able to duplicate the anomaly on the SUN by running the utility "ansitape.c" while omitting the carriage control parameter (cc=e). The same "ansitape.c" utility was then used to read that tape back to disk . The resulting disk files were flawed with "line-feed" characters in position 129 of each record. This tape, like the original Autotrol tape, was acceptable to the VAX but rejected by the SUN.

Some subtile difference existed in the format structure (not the data) of ANSI tapes generated (by "ansitape.c") with the "cc=e" switch on. More over, this difference was distinguishable by the SUN but transparent to the VAX.

ANOMALY DETECTION and DEFINITION

Two almost identical tapes were made on the SUN. Using the same source files (un-flawed Group-4 images), one tape was written with the carriage control parameter and the other tape was written without:

> ansitape -rv mt=/dev/rmt8 rf=f rs=128 d001r001
> ansitape -rv mt=/dev/rmt8 fr=f rs=128 cc=e d001r001

Both tapes were acceptable to the VAX while only the latter could provide images successfully displayable by the SUN. After dumping the raw format structure of both tapes, it was discovered that a format control flag in an ANSI header was missing from the tapes producing the anomaly. The flag is in byte 27 of the "Reserved for Systems Use" field of ANSI header 2 (HDR2).

This flag is documented in the DEC VAX/VMS Guide to Magnetic Tape and Disk Operations and in the US Army Artificial Intelligence Center documentation for the public domain utility "ansitape.c".

Apparently, various system environments deal with this flag differently. The system on which the data was originally generated (an APOLLO) disregards the flag completely. The CTN VAX writes the flag when generating a raster file but ignores the flag while reading a raster file. The CTN SUN system (using "ansitape.c") will write the flag when it generates a new raster file and requires the flag when it reads a raster file.

Several CTN participants were poled to ascertain their implementation of the ANSI tape standard:

> One found the same problem transferring ANSI tape data between their Apollo system and their Sun system. Autotrol uses public domain software ("ansitape.c") to read and write ANSI formatted tapes on the Sun. Their solution was to modify "ansitape.c" to allow a parameter in the read command to strip or add line-feed characters on input, as required.
>
> Two others were unaware of the carriage-control flag in HDR2 but had run into the effect of the anomaly before.

The CTN VAX is unaffected by this anomaly since it ignores the carriage control flag during data input, allowing it to successfully read files regardless of the flag. Further, the CTN VAX is able to transfer data out successfully because it implements the flag during a tape write.

The Army Artificial Intelligence Center, who produced "ansitape.c", was not aware of the anomaly since they used that utility exclusively on ASCII/EBCDIC data transfers. Additional information on the development of "ansitape.c" was not available because the individual responsible no longer worked for the Army.

STANDARDS ISSUES

In order to achieve successful raster data transfer between various systems, a uniform data definition must be adopted. The definition must apply to all levels of the interchange mechanism, media, media format structure, file structure and record structures within the files.

Fixed length records are established on a system as contiguous strings of bytes, each string having a "fixed length" as specified. No extra record delimiters are required between records.

Within the various standards called out in the CALS initiative, all levels of data interchange are specified. The carriage-control flag causing the anomaly, while not specified by the standard, is not explicitly excluded from it.

The files generated by a system reading CALS tapes should be viewed as part of a particular application. It is up to the implementor to reconcile the disk file format of the image files captured from CALS interchange media by the host system.

The following are excerpts from the standards and may provide some insight as to how file management strategies and attributes may affect data file interpretation when transferring data between dissimilar systems.

> MIL-STD-1840A specifies that ANSI X3.27 provide the standard for data interchange. 1840A further specifies:
>
> "5.2.1.6 Raster Files: The data in the first block of a raster file shall be written with 128 byte ANSI fixed length records...."

ANSI X3.27-1978 specifies magnetic tape label, file and record structure for data interchange. It assumes that various types of file structures exist but does not require any conformity in these structures:

> "1.3 EXCLUSIONS
>
> 1.3.1 This standard assumes the existence of but does not specify:
>
> (7) Common programming language declarations and processing statements to define the attributes of output files, to identify input files, to process user file labels, and to process records within the file.
>
> (8) A method for an originator to communicate with a recipient, a description of a file, including, but not limited to:
>
> > (c) The content , layout, and format of the fields in each record type.
> >
> > (d) The indicator of each of the record types.
> >
> > (e) The key for each of the records.
> >
> > (f) The structure or sequence of the records in the file."

Also:

"7.6 Second File-Header Label (HDR2)

7.6.4 Reserved for System Use (CP 16-50) On output or input, or both, this field may be used by a system that recognizes System Code (HDR1 CP 61-73). In interchange the contents of this field are ignored."

ANSI X3.27 allows various file and record structures to exist and even provides fields for the interchange of system specific control data.

The ANSI tape standard is intended to interchange magnetic tape recorded data between similar and potentially different computer system architectures. To accommodate this capability the format includes a wide range of attributes indicated in the ANSI file headers. Among these is a "System Code" (re. ANSI X3.27 4.5) in bytes 61 to 73 of each tape file header-1 ("HDR1"). This is an ASCII field that identifies the system that recorded the data.

The format of the identifier is not specified by the ANSI standard but is left up to the system's implementor. In this manner, a system receiving an ANSI X3.27 formatted tape is able to determine if the data being interchanged is from a known or an unknown source. With such information, receiving systems know whether to apply or ignore other ANSI header information such as the "Reserved for System Use" filed in each tape file header-2 ("HDR2"). This is an ASCII field intended for system specific information, it is not intended for use during data interchanges between dissimilar systems.

The concepts of "file" and "record" (to some extent) are system specific, and are embodied in the design of the system. ANSI X3.27 is used to define fixed- and variable- length records (re. ANSI X3.27 62.2. and 6.2.3) for the purpose of CALS data interchanges.

Issues, such as systems' conventions with respect to imposing carriage control characters on input or skipping over or selectively processing the circumflex accent, are not universal.

RECOMMENDATION

The "Carriage Control" flag though documented in parochial literature (DEC and "ansitape.c") does not appear to have a consensus of use driving it.

The specified standard ANSI X3.27 (Magnetic tape labels and file structures for information interchange) does not specify this data item.

Locations 16 to 50 of tape "HDR2" are specified as "Reserved for System Use" and "not intended for use in an interchange environment".

Since the Specification does not directly disallow the use of the data flag or the use of the header area it is being transferred in, the use of the flag by some systems can not be prevented.

Conversely, the need for some systems and/or utilities for this flag is not explicitly sanctioned by the specification and therefore can not be required.

The position of the CTN is that any data may exist in tape "HDR2", bytes 16 to 50, but this data shall not be required to allow a system to correctly interpret a MIL-STD-1840A data set. CALS implementations using utilities such as "ansitape.c" may require source code or procedural

modification to successfully read ANSI formatted tapes that pass no information in tape "HDR2", bytes 16 to 50.

A warning should be posted to CTN participants, that the use of "ansitape.c", without modification, may produce unusable Group-4 files on magnetic disk by reading tapes which do not have an "M" in location 27 of the "Reserved for Systems Use" filed in "HDR2".